Rapid Data Assimilation in the Indoor Environment: theory and examples from real-time interpretation of indoor plumes of airborne chemicals

Ashok Gadgil, Michael Sohn, Priya Sreedharan

Environmental Energy Technologies Division
Lawrence Berkeley National Laboratory
1 Cyclotron Road
Berkeley, CA 94720 USA

September 2008

This work was supported by the Office of Chemical Biological Countermeasures, of the Science and Technology Directorate of the Department of Homeland Security, and performed under U.S. Department of Energy Contract No. DE-AC02-05CH11231

Keynote paper Topic 3 “Data Assimilation”.
IMT29, Aveiro, Portugal. September 24-28, 2007

Rapid Data Assimilation in the Indoor Environment: theory and examples from real-time interpretation of indoor plumes of airborne chemicals

Ashok Gadgil
Senior Staff Scientist and Group Leader
Lawrence Berkeley National Laboratory
1 Cyclotron Road, Berkeley, CA 94720 USA
and
Adjunct Professor, Energy and Resources,
University of California, Berkeley, CA 94720 USA

E-mail: ajgadgil@lbl.gov
Phone: +1-510-486-4651
Fax: +1-510-486-6658

Abstract:
Releases of acutely toxic airborne contaminants in or near a building can lead to significant human exposures unless prompt response measures are identified and implemented. Commonly, possible responses include conflicting strategies, such as shutting the ventilation system off versus running it in a purge (100% outside air) mode, or having occupants evacuate versus sheltering in place. The right choice depends in part on quickly identifying the source locations, the amounts released, and the likely future dispersion routes of the pollutants. This paper summarizes recent developments to provide such estimates in real time using an approach called Bayesian Monte Carlo updating. This approach rapidly interprets measurements of airborne pollutant concentrations from multiple sensors placed in the building and computes best estimates and uncertainties of the release conditions. The algorithm is fast, capable of continuously updating the estimates as measurements stream in from sensors. The approach is employed, as illustration, to conduct two specific investigations under different situations.

1.0 INTRODUCTION AND MOTIVATION

Airborne acutely toxic contaminant releases in or near a building can lead to significant human exposures unless prompt response measures are taken. However, possible responses can include conflicting strategies, such as shutting the ventilation system off versus running it in a purge mode, or having occupants evacuate versus sheltering in place. The right choice depends

in part on knowing the source locations, the amounts released, and the likely future dispersion routes of the pollutants. Determining this information is complicated by the complex nature of airflows typically found in multi-room, multi-story buildings. Merely detecting an airborne pollutant from one or more sensors placed in the building, may not reveal the location or strength of the source. The sensor measurements must be interpreted to estimate the source characteristics and quantify the uncertainties. For effective decision making, the measurements must also be interpreted quickly and continuously as data stream in from the sensors.

Traditional algorithms for data interpretation generally use an inverse modeling approach (e.g., optimization and Gibbs sampling) to fit an indoor airflow and pollutant transport model to measurements of airborne pollutants. The fit is usually achieved by iteratively adjusting model input parameters until they reasonably predict the data. For on-line, real-time, sensor data interpretation, these approaches are too slow. They (i) wait to execute computationally intensive fate and transport models until data are first obtained; (ii) execute the models repeatedly as new or successive sensor data become available; and (iii) require a considerable amount of data before the algorithm finds a unique solution or estimates the uncertainty in the calibrated parameters. Lastly, the computational burdens required by these algorithms can be so great that using them for pre-event planning, such as to determine optimal monitoring locations, sampling plans and sensor performance criteria, can be impractically cumbersome.

Many of these problems can be solved using a technique called Kalman filtering [see e.g., 1]. It is well-suited for many sensor interpretation applications and has been successfully applied, for example, to estimate the source strength of pollutant releases in multizone buildings [2]. However, it is best used for linear systems with well-conditioned input-to-output parameter covariance matrices and strong observability between the internal state variables (e.g., the model input parameters of an indoor airflow and pollutant transport model) and the model outputs (e.g., concentration predictions) [1, pg. 209 and 281]. Many fate and transport phenomena, such as second-order pollutant degradation, density-driven pollutant transport, aerosol coagulation, and second-order pollutant diffusion in sorption-desorption, are not linear. Furthermore, there exist wide uncertainty bounds for several of the model inputs, such as many possible source locations, amounts released, durations of releases, and HVAC operating conditions. These will invariably lead to ill-conditioned covariance matrices and poorly observable systems. Though many nonlinear models may be linearized using an extended Kalman filtering technique [3], the technique requires considerable tuning and adjustments due to the linear approximations. Lastly, Kalman filtering is not well-suited for the uncertainty analyses required for effective decision analysis. Kalman filtering may be applied to estimate the most likely source location and uncertainty (for example, “the source is in room ‘A’ with 80% probability”). It is however considerably more computationally-intensive, and thus time-consuming, to concurrently estimate the uncertainties of other less likely source locations (for example, “the source is in room ‘A’ with 80% probability, in room ‘B’ with 18% probability, and in room ‘C’ with 2% probability).

We describe an alternative algorithm which uses Bayesian statistics. This approach succeeds where traditional methods fail because it (1) decouples the simulation of predictive fate and transport models from the interpretation of measurements and (2) incorporates uncertainty analysis in all parts of the framework. Thus, we can compute the time-consuming airflow and pollutant transport predictions and uncertainty estimates – without requirements on linearity of the models – prior to a pollutant release event, and interpret sensor data in real time during an event. The technique may be used to estimate the location, magnitude, and duration of the release, to characterize any unknown or variable building or weather conditions, and to predict

future pollutant transport in the building. Initial estimates are provided as soon as a sensor detects a pollutant, and can be updated as each new measurement arrives.

The techniques we introduce are not new; often termed Bayes Monte Carlo updating, it has been applied to assess environmental health risk [e.g., 4, 5, 6], analyze groundwater monitoring data [e.g., 7, 8, 9], and conduct environmental value-of-information analyses [e.g., 10, 11]. However, the research problems described in these papers are distinct from the current work in one important feature. The papers describe applications to interpret data well after they were collected, when interpretation and response were not time-critical. In the present work, we recognize and exploit a feature of Bayes Monte Carlo updating not previously recognized: modeling and data analysis can be decoupled, which allows for data to be interpreted while they stream in during a pollutant release event. This is a significant advance over previous uses of Bayes Monte Carlo methods. Furthermore, our research group has been the leader in application of this general approach to indoor air pollutant source characterization and airborne pollutant transport predictions.

In this keynote paper, we (i) elucidate the Bayesian algorithm for interpreting sensor data in real time, and (ii) demonstrate the approach with two selected examples.

In the first example, we successfully detect and characterize a pollutant release in a hypothetical five-room building. In this illustrative application, we generated synthetic data for two data collection scenarios: (1) concurrent sampling, in which sensor measurements are obtained simultaneously in each of the five rooms at five-minute intervals; and (2) sequential sampling, in which sensor measurements are obtained sequentially, one room at a time, at five-minute intervals. We also generated high, medium, and low quality data for each of the two data collection scenarios to examine degradation of the predictive results with increasingly noisy data. In addition to unknowns regarding the location, duration, and magnitude of the pollutant release, other unknowns included whether certain doors or windows were open or closed, and the outside temperature. Lastly, interpreting data for several sampling plans and qualities of data demonstrates the ease of exploring and comparing the trade-offs between sensor features, for example, frequency of sampling, sensor sensitivity, and number of sensors.

2.0 APPROACH

Since this is a diverse audience, we take the time to explain the approach assuming that not all readers are familiar with Bayesian Monte Carlo updating.

The Bayesian data interpretation approach is divided into two stages. First, in the *pre-event* or *simulation* stage, the practitioner selects a fate and transport model, builds a computer model of the building, characterizes uncertainties of the model inputs, and simulates many hypothetical airflow and pollutant transport scenarios. These time-consuming tasks are completed before a pollutant release occurs. In the second stage, during a pollutant release event, the agreement between each of the model simulations and sensor data are evaluated using a technique called Bayesian updating [see e.g., 5, 9]. This stage is quick and is conducted as data stream in from the sensors.

2.1 Pre-Event Planning

Before a release event, the practitioner develops a model of the building's indoor airflow and pollutant transport. Best estimates for model inputs are generated from, for example, previous building characterization exercises, tracer gas flow experiments and modeling, published literature, and professional judgment. Any uncertain model input parameter or variable building characteristic is assigned an uncertainty distribution that describes the probabilistic range of possible values. Pollutant description uncertainties, e.g., the location,

duration, and amount of pollutant released in an incident, are also assigned uncertainty distributions. Generally, wide distributions are assigned due to the limited prior information, particularly for describing the pollutant characteristics.

The practitioner next generates a library of model simulations by sampling the distributions of the model input parameters using Monte Carlo or other sampling technique, and predicting airflow and pollutant transport for each set of parameters. Each model simulation represents a possible building configuration and pollutant release scenario. At this stage, each simulation is equally likely to occur. Thus, sufficient sampling of the uncertainty distributions is essential to represent the full range of possible building and pollutant release characteristics. One method for testing sufficiency of sampling is by increasing the sample size until changes in summary statistics (e.g., means, variances, coefficients of variation) of model predictions are negligible. The resulting library of simulations may consist of several thousand scenarios. Since this stage is not time critical, a large library of simulations is not difficult to develop with the advances of fast personal computers and inexpensive data storage devices.

It is important that the parameter ranges sampled in the uncertainty characterization and Monte Carlo sampling are wide enough to contain the parameter values of the actual event (to be diagnosed in real-time). Otherwise, the method will fail to converge on the correct parameter values. Of course such failure still provides some useful information: that the actual event parameters are not within the ranges sampled, or there is a model-misfit to the actual event.

2.2 During Event Data Interpretation

During a release event, the algorithm compares data streaming in from sensors to each realization in the library of model simulations using a structured probabilistic method referred to as Bayesian updating. Bayes' rule allows the practitioner to quickly estimate and update the level of agreement between the observed data and each model simulation (i.e., the pollutant transport predictions). To summarize the process, each realization in the library is compared to the data to quantitatively assess the likelihood that the realization describes the event in progress. A high likelihood estimate will result for the realization(s) (i.e., the model simulation(s) with predictions) that fit the sensor data well. This in turn suggests that the model inputs used to generate that realization in the pre-event simulation stage have high probability of describing the event in progress. By rapidly evaluating and comparing the relative fits for each realization using Bayesian statistics, the practitioner estimates the best-fitting suite of model inputs and the associated uncertainty.

The difference between the data and the predictions resulting from measurement error, spatial and temporal averaging or correlations, and imperfect model representation are all considered when quantitatively estimating the data-to-model agreement. The probability of each model simulation before and after assessing the agreement is termed the *prior* and *posterior* probability, respectively.

Suppose the library of Monte Carlo simulations contains “N” simulations (or realizations), and each one is indexed with a number, “*k*”. With each realization, there is associated a set of output parameters (as a functions of time), and also input parameter values. “Updating” involves updating the probability assigned to each of the N realizations in the library by comparing the freshly obtained data (say, at a given time) with the corresponding prediction for that time by each of these N realizations. Before the updating, their probabilities are called “prior probabilities” and the updating process results in the “posterior probability” after that freshly obtained data is assimilated into the assigned probability for each realization.

The posterior probability of the k^{th} Monte Carlo simulation making prediction Y_k given the sensor measurements O is denoted as $p(Y_k|O)$. The prior probability of the same k^{th} Monte Carlo simulation making prediction Y_k is denoted as $p(Y_k)$. Before onset of data comparison, each of the model realizations are usually assumed equally likely (i.e., $p(Y_k) = 1/K$). As data streams in, for each time step, we compare the observations, denoted by "O" , with predictions "Y_k" made by each of the N realizations, where the subscript k refers to the predictions for that time made by the k^{th} realization.

Using Bayes' rule, $p(Y_k|O)$ is calculated using eq 1 [5]:

$$p(Y_k \mid O) = \frac{p(O \mid Y_k)p(Y_k)}{\Sigma_{i=1}^{N} p(O \mid Y_i)p(Y_i)} \tag{1}$$

where $p(Y_k|O)$ is the posterior probability; $p(O|Y_k)$ is the likelihood of observing measurements O given model prediction Y_k; $p(Y_k)$ is the prior probability of the k^{th} Monte Carlo simulation; and N is the number of Monte Carlo simulations. The posterior probability, $p(Y_k|O)$, describes the probability of all of the model assumptions and predictions associated with the k^{th} realization. Thus, the prior uncertainty of each model input parameter (e.g., source location or building characteristic) and model output (e.g., airborne concentration prediction) is updated according to how well model predictions in the prior uncertainty distribution agree with the sensor data.

The posterior probability (Bayes Factor) of each simulation is applied to each parameter associated with that simulation, and thus a re-estimation is obtained for all input uncertainties. The posterior mean is calculated for each input parameter and a weighted estimate of the value of the parameter is obatained [5].

(2)

The likelihood function, $p(O|Y_k)$, in eq 1 quantifies the error structure of the data, i.e., the differences between the data and the model predictions resulting from measurement error, spatial and temporal averaging or correlations, and imperfect model representation. If "S" independent measurements are considered, for example following sequential concentration measurements

$$Mean: \quad \mu_V' = \sum_{i=1}^{U} V_i * p_i'$$

returned from sensors or from concurrent measurements sampled in several locations, the likelihood of observing all of the measurements is the product of all of the individual likelihoods:

$$p(O \mid Y_k) = \prod_{j=1}^{S} p(O_j \mid Y_k) \tag{3}$$

For unbiased measurements with a normally distributed error, the likelihood of observing a sensor measurement, O_s, given a model prediction, $Y_{s,k}$, is given as:

$$p(O_s \mid Y_k) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{O_s - Y_k}{\sigma_\varepsilon} \right]^2 \right)$$

(4)

where O_s is the concentration measured by a sensor in a room at $t=t_s$; $Y_{s,k}$ is the airborne concentration predicted from the k^{th} Monte Carlo realization that corresponds to O_s; and σ_ε^2 is the error variance of the measurements. The error variance, σ_ε^2, describes not only the error in the sensor instruments, but also the error associated with comparing model predictions with sensor measurements having different spatial and temporal averaging. Alternative likelihood function can be used, as appropriate, without affecting the overall approach.

This second stage of the approach is mathematically simple and can be executed very quickly, much quicker than the rate at which new data are likely to arrive from sensors.

3.0 FIRST ILLUSTRATIVE APPLICATION

This first illustration is to locate and characterize a hypothetical pollutant release in a five-room building. Uncertainties in source location, duration, and amount, and in some building characteristics, were estimated and updated using synthetic data.

3.1 Building Description

The study building is a single story building comprising three rooms, a common area (CA), and a bathroom (Figure 1). Each of the partitioned areas are treated as well-mixed zones. The zones connect to the outside via windows and doors, and interconnect via internal doors. The building does not have an HVAC system.

The status of one of the CA zone windows and the door between the CA zone and Room 3 is "unknown" to the simulations (e.g., owing to failed position sensors at these locations). These are denoted in Figure 1 with question marks. All other windows and doors are closed. Wind blows at a steady 3 m/s on the exterior wall shared by the CA zone and Room 1. The temperatures of the rooms are indicated in Figure 1 and the outside temperature is unknown.

Details of the uncertainties in the source and building characteristics will be found in Sohn et al [12].

3.2 Airflow and Pollutant Transport Simulation

We selected the COMIS model [12] to predict indoor airflow and pollutant transport. COMIS predicts the steady-state flows of air and the dynamic transport of pollutants by representing the building as a collection of well-mixed zones. Air flows between zones via cracks, doors, and windows (and also fans and ductwork, although we did not use these features here). Though we selected a multi-zone modeling approach for this application, our data interpretation algorithm may be used with any suitable indoor airflow and pollutant transport model.

As part of the pre-event planning, we generated a library of airflow and pollutant transport simulations by sampling the uncertainty distributions describing the source and building characteristics. Five thousand air flow and pollutant simulations, each of them equally likely, were generated using Latin Hypercube sampling techniques [13]. Means and variances for several sample sizes were tested to ensure that five thousand simulations adequately sampled the problem solution space.

3.3 Description of Synthetic Data

We generated synthetic data to represent measurements that might stream in from air monitoring sensors placed in the building. The synthetic data are based on an airflow and pollutant transport simulation that represents a possible pollutant release event. Figure 2 plots the simulation from which the synthetic data were generated and Table 3 summarizes the model

input information. The simulation was excluded from the library of five thousand simulations describing the prior pollutant concentration predictions.

We added measurement error to the model simulation using a two-part error structure: (1) a normally distributed error associated with a standard deviation proportional to the true value, i.e., a fixed coefficient of variation, and (2) a normally distributed random error that is independent of the magnitude of the measurement. We are not limited to this structure for errors. Many statistical methods for handling more complex error structures are available and can be used in place of eq. 4 [see e.g., 14, 9].

We generated high, medium, and low quality synthetic data with progressively larger magnitudes of error components in the error structure. If adding the error generated a negative value for the pollutant concentration, we set the simulated measurement to zero. Details are given in [12], and Figure 2 shows the synthetic data in the CA zone and Room 1. As expected, the high quality data (Figure 2a) shows a more consistent pollutant concentration time series than is described by the low quality data (Figure 2c).

3.4 Data Interpretation

Along with the three levels of data quality (Section 3.3), we also evaluated two different plans for data collection. In the first, *concurrent sampling*, we obtain sensor measurements from all five zones simultaneously at five-minute intervals. In the second, *sequential sampling*, we obtain sensor measurements sequentially, one zone at a time, at five-minute intervals. Though we present the results for the various quality of data and sampling schemes, it is important to emphasize that the results merely illustrate the types of data interpretation and "what-if" analyses that may be conducted using our interpretation algorithm. The results do not represent the success or failure of the interpretation approach.

Data interpretation

Figure 3 shows the estimation of the source location for the three qualities of data. With medium or high quality data, the interpretation correctly identified the source location at the first measurement event (t=5 min.), when five measurements were obtained. With low quality data, the identification of the source location was slower, requiring more measurements to overcome the error in the data. More details are given in [12]. Again, the medium and high quality data permit dramatic uncertainty reductions at t=5 min., in all cases converging to the correct answers. The low quality data however required more data, and thus more time.

With the sequential sampling plan, we gather data at a significantly slower pace – one, not five, data points every five minutes. It might represent a situation where a single (expensive) sensor is multiplexed to several sampling tubes. The rooms were sampled in the order (1) CA zone at t=5 min., (2) Room 1 at t=10 min., (3) Room 2 at t=15 min., (4) bathroom at t=20 min., (5) Room 3 at t=25 min., and (6) CA zone at t=30 min. Figure 3 shows the estimation of the source location. Bearing in mind that the sequential sampling plan collects five times less data than the concurrent sampling plan, the medium and high quality data did not locate the source until all of the rooms were sampled once (t=25 min.), though reasonably good estimates were generated as early on as t=10 min. The low quality data, however, did not locate the source even after 30 min.

4.0 SECOND ILLUSTRATIVE APPLICATION

4.1 Problem Description

Now we focus on trigger- or alarm-type sensors, rather than continuous-output devices. We base our case study on data from one of twelve tracer-gas experiments conducted at a three-

story, 660 m^3 building at the Dugway Proving Ground, Utah [15]. The paper from which this these illustrations are taken, [16], examines how well various sensor systems, each system consisting of sensors with different sensor characteristics (threshold level, response time, and accuracy), reconstruct the release event. These examples in [16] demonstrate the importance of a systems perspective in selecting sensors with desirable sensor characteristics. However, for lack of space, only one illustrative example will be shown here.

4.2 Approach

We consider the following problem. A contaminant is released, over a short duration, somewhere in a building (this may include its indoor air intake vents). A network of threshold or alarm-type sensors operates to detect the contaminant. We seek to understand how sensor characteristics such as threshold level and response time affect the ability of a sensor interpretation algorithm to quickly detect and characterize the contaminant release.

For the purposes of this paper, we assumed that each zone is equipped with one sensor, and each sensor has a single threshold, meaning the output is either "below threshold" or "above threshold." We define several possible threshold levels that are at or above the minimum detection limit of the sensor. Three important parameters characterize sensor performance in this study: threshold level; response time (also sometimes called integration time); and accuracy.

A sufficiently large release will trigger one or more sensors. A Bayes Monte Carlo algorithm is then initiated to determine key information about the release event, including the location of the release and the contaminant mass discharged. This information can help guide emergency response and post-event remediation.

4.3. Building characterization, data collection and model generation

The study focuses on one unit in a multi-unit building located at the Dugway Proving Grounds, Utah [15]. The unit consists of 660 m^3 of interior volume and approximately 280 m^2 of floor area on three levels. A mechanical air-handling unit (AHU) supplies air to the first and second floors, and its return is on the first floor. The AHU is a 100% recirculating unit (i.e., there is no deliberate outside air intake).

A library of 5000 simulated contaminant releases was generated using a COMIS model of this building, which was based on detailed measurements of the building flow paths. For the library, we sampled from statistical distributions of a set of key input parameters, as described earlier.

Tracer gas dispersion data for examining the interpretation of signals streaming from triggered sensors was obtained from tracer gas experiments conducted in the same building, and described in [15].

We generated hypothetical threshold sensor data by interpreting the tracer data from the experiment [15] as if they were concentrations to which surface acoustic wave (SAW) sensors were exposed. SAW sensors are piezoelectric devices, often configured to provide alarms based on whether the incoming concentrations are above or below a predefined trigger or threshold level. False positive and false negative alarms may occur, according to the performance characteristics of each sensor, or the inability of the sensor to distinguish between the contaminant of interest and interfering chemicals that also may be present in the air.

We generated hypothetical alarm data for sensors with different performance characteristics, based on discussions with several developers and users of SAW sensors. Three sensor attributes were varied: threshold level, response time, and error.

The threshold levels were chosen relative to the measured concentrations during the first 120 minutes of the release event. The lowest selected threshold would cause 98% of the data to trigger the alarm, while the highest threshold would trigger an alarm for only 1% of the data. However, for presentation purposes, we normalize both threshold levels and concentration data by the concentration that would be found in the release zone if the entire release amount instantaneously mixed throughout that zone. That is, thresholds and concentrations are reported in terms of the theoretical maximum peak concentration that could be measured in the system under the perfectly well-mixed assumption. With this normalization, the lowest threshold level was 0.02% of the maximum peak, and the highest was 16%.

Sensor response times ranged from 20 seconds to 180 seconds. In the simulations, concentrations are averaged over the response time, and then compared to the appropriate threshold level. Note that averaging over the response time corresponds to an assumption that the SAW desorption cycle is brief relative to its adsorption cycle. In our simulations we ignored the duration of the desorption cycle (i.e., each sensor started integrating the next cycle of data as soon as it reported an alarm or no-alarm condition).

Simulations were run using data with and without synthetically added error. For simulations with added error, we generated sensor signals according to the following assumptions: (1) if the actual concentrations are within 25% of the sensor threshold level, the signal will be false 50% of the time; and (2) for concentrations outside of this range, the signal will be false either 10% or 30% of the time, depending on the assumed sensor error.

In this implementation of Bayes' rule, the likelihood function is based on the probability used to generate the false positives and negatives. For example, for data generated with a 30% error, the likelihood is 0.3 when the modeled concentration is more than 25% above the threshold level and the sensor has not signaled on; conversely, the likelihood is 0.7 if the sensor has signaled on. For the simulations using data without synthetically added error, we assume 5% error for all data. We did not assign 100% confidence to this data because of inherent uncertainty and variability. In practice the designer of the sensor system should have reliable information on the sensor's actual rate of false positives and false negatives.

Figure 4 illustrates the conversion of measured concentration data to simulated threshold data. Figure 4(a) shows normalized time-averaged concentration data, with the threshold level indicated. Figure 4(b) shows the threshold data that would result from a threshold sensor with no error and instantaneous response. Here, "1" signals that the concentration exceeds the threshold. Figure 4(c) shows the threshold data, corrupted with false negatives or positives. Because the false readings are generated stochastically, different realizations of the data in Figure 3(c) would exhibit different patterns of output signals.

For the case-study examples, we systematically varied the sensor characteristics, as described above. In cases where error was specifically investigated, for each sensor attribute, we generated 50 sets of error-added threshold data, analogous to those displayed in Figure 4(c), for each sensor in the system. Each combination of threshold level, response time, and error produced a data stream against which to apply the Bayes Monte Carlo algorithm. The algorithm was used to determine the release location and release magnitude; the time of release was assumed known.

4.4 RESULTS FOR TRIGGERED SENSORS

To demonstrate data interpretation using threshold sensor data, we limit this paper to only showing the ability of the sensor system to estimate the release location. Interested readers are referred to [16] for more results and discussion.

The information content in threshold sensor data is significantly less than that in direct concentration measurements. Nevertheless, the sensor system can successfully reconstruct the source, at least in some circumstances. We demonstrate this with an example in which the concentration data have been converted to threshold data using a threshold level of 2.3%, a sensor response time of 20 seconds, and without added error. We judge the sensor system performance by its ability to reduce the uncertainty in the estimates of the release location, mass, and duration parameters, and by the time required to do so.

Figure 5 depicts the time required to identify the release location (Room 1.2a). At time zero, every zone is assumed to be equally likely as the release location. As sensor data arrive, the Bayes algorithm adjusts these probabilities, locating the release location with greater than 90% confidence within one minute. If rapid response hinges on locating a source very quickly, this example suggests that threshold sensors under this network configuration and data quality may be acceptable for real-time monitoring.

5. CONCLUSION

Real-time environmental monitoring systems have the potential to help protect building occupants in the event of high-risk pollutant releases. Selection of sensor characteristics is best performed from a systems perspective. Here, we have demonstrated — albeit for a limited set of circumstances — that a network of continuous or single-level threshold sensors can be used to determine the location and magnitude of the release within a Bayes Monte Carlo framework. More importantly, treating the network as a system may lead to better choices for sensor characteristics like response time and error, than might be the case when considering sensors individually.

With more complex buildings, system characterization is technically more challenging, and also more expensive. Hybrid methods that augment prior knowledge with sensor system data that monitors building operations to learn about airflows and contaminant transport may improve overall system performance. Such advances would not only be beneficial for designing indoor monitoring systems, but may potentially be extended to help diagnose and interpret data from large-scale contaminant releases to the ambient atmosphere and to other environmental media. Such approaches also hold the promise of facilitating improvements in building performance with respect to energy use, thermal comfort, and indoor air quality.

ACKNOWLEDGEMENTS

This work was supported by the Office of Chemical Biological Countermeasures, of the Science and Technology Directorate of the Department of Homeland Security, and performed under U.S. Department of Energy Contract No. DE-AC02-05CH11231

REFERENCES

1. Grewal, M.S.; Andrews, A.P. *Kalman Filtering: Theory and Practice*; John Wiley & Sons: New York, 2001; 2nd edition.
2. Federspiel, C.C. Estimating the Inputs of Gas Transport Processes in Buildings; *IEEE Transactions.* **1997**, 5(5), 480-489.
3. Del Gobbo D; Napolitano M; Famouri P; Innocenti M. Experimental Application of Extended Kalman Filtering for Sensor Validation; *IEEE Trans. on Control Systems Technology*. **2001**; 9(2), 376-380.
4. Taylor A.C.; Evans J.S.; McKone T.E. The Value of Animal Test Information in Environmental Control Decisions; *Risk Analysis.* **1993**; 13(4), 403-412.
5. Brand K.P.; Small M.J. Updating Uncertainty in an Integrated Risk Assessment: Conceptual Framework and Methods; *Risk Analysis.* **1995**; 15(6), 719-731.
6. Pinsky P.F.; Lorber M.N. A Model to Evaluate Past Exposure to 2,3,7,8,-TCDD; *J. Exp. Anal. and Env. Epid.* **1998**; 8(2), 187-206.
7. Dilks D.W.; Canale R.P.; Meier P.G. Development of Bayesian Monte-Carlo Techniques for Water Quality Model Uncertainty; *Ecol. Model.* **1992**; 62, 149-162.
8. Wolfson L.J.; Kadane J.B.; Small M.J. Bayesian Environmental Policy Decisions: Two Case Studies; *Ecological Applications.* **1996**; 6(4), 1056-1066.
9. Sohn M.D.; Small M.J.; Pantazidou M. Reducing Uncertainty in Site Characterization Using Bayes Monte Carlo Methods; *J. of Env. Engineering.* **2000**; 126(10), 893-92.
10. Finkel A.M.; Evans J.S. Evaluating the Benefits of Uncertainty Reduction in Environmental Health Risk Assessment; *JAPCA.* **1987**; 37, 1164-1171.
11. Dakins M.E.; Toll J.E.; Small M.J.; Brand K.P. Risk-Based Environmental Remediation: Bayesian Monte Carlo Analysis and the Expected Value of Sample Information; *Risk Analysis.* **1996**; 16(1), 67-79.
12. Sohn, M. D., Reynolds, P., Singh, N., and Gadgil, A. J. Rapidly Locating and Characterizing Pollutant Releases in Buildings. Journal of Air and Waste Management Association, **52:** pp. 1422-1432.
13. Iman R.L.; Conover W.J. Small Sample Sensitivity Analysis Techniques for Computer Models, with an Application to Risk Assessment; *Commun. Statist. - Theor. Meth.* **1980**; A9(17), 1749-1842.
14. Morgan M.G.; Henrion M. *Uncertainty, A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*; Cambridge University Press: New York, 1990.
15. Sextro, R.G., Daisey, J.M., Feustel, H.E., Dickerhoff, D.J., Jump, C., 1999. Comparison of modeled and measured tracer gas concentrations in a multizone building. Proceedings of the 8th International Conference on Indoor Air Quality and Climate – Indoor Air 99, Vol 1, pp 696-701. Indoor Air 99, Edinburgh.
16. Sreedharan, P., Sohn, M. D., Gadgil, A. J., and Nazaroff, W. W. Systems approach to evaluating sensor characteristics for real-time monitoring of high-risk indoor contaminant releases. Atmospheric Environment **40,** (2006) pp. 3490-3502.

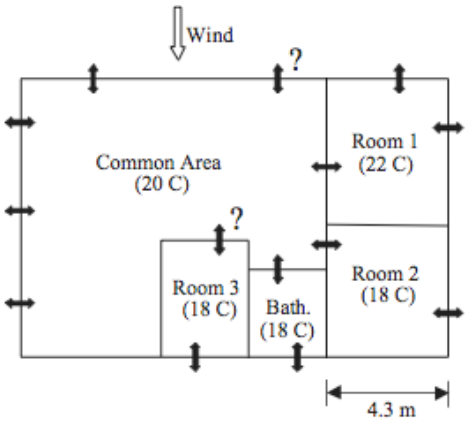

Figure 1. Plan of the five-room building. The arrows represent windows or doors; the question marks indicate uncertain open or closed status. The windows in the bathroom and Room 2, and interior door for Room 1 are open. All other windows and doors are closed. The wind blows at a steady 3 m/s.

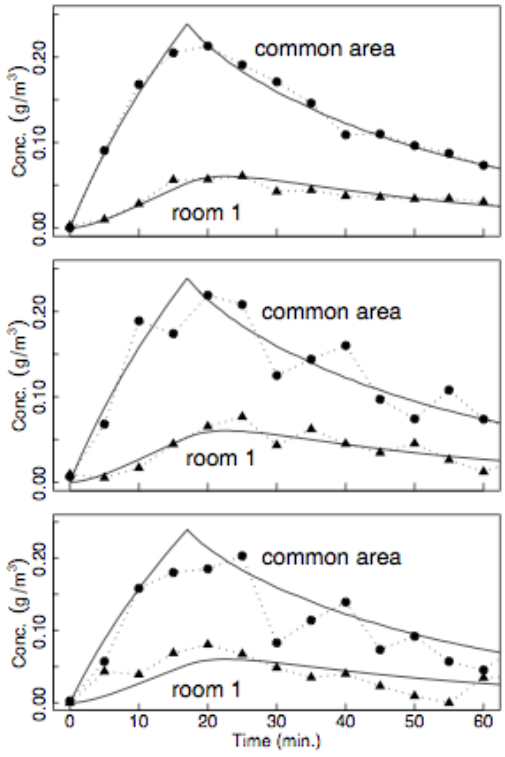

Figure 2. Synthetic data representing (a) high, (b) medium, and (c) low quality measurements in the common area zone and Room 1. The solid lines represent the model simulation originally used to generate the data. This particular instance of the model realization was removed from the Monte Carlo library before the data interpretation was started.

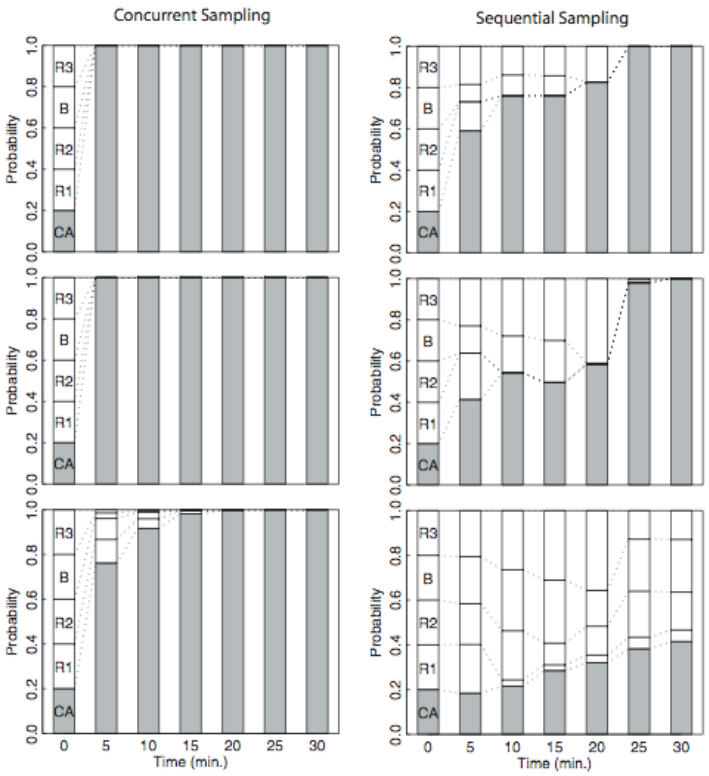

Figure 3. Locating the source using (a) high, (b) medium, and (c) low quality measurements. Concurrent sampling draws a measurement from each zone every five minutes. Sequential sampling draws a measurement from one room at a time every five minutes in the order: (1) common area at t=5 min., (2) Room 1 at t=10 min., (3) Room 2 at t=15 min., (4) bathroom at t=20 min., (5) Room 3 at t=25 min., and (6) common area at t=30 min. The probabilities at t=0 is before data interpretation.

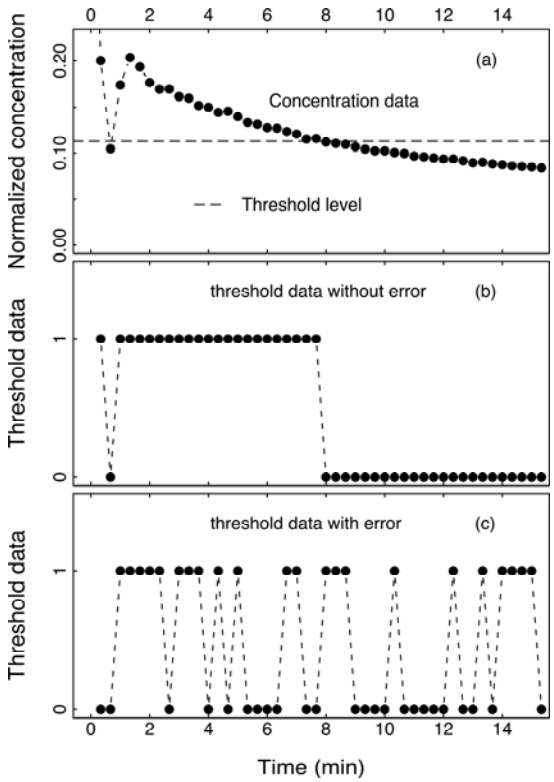

Figure 4. Sample illustrating conversion of tracer gas concentration to threshold data: (a) concentration data; (b) threshold data without simulated error added; (c) threshold data with simulated error added.

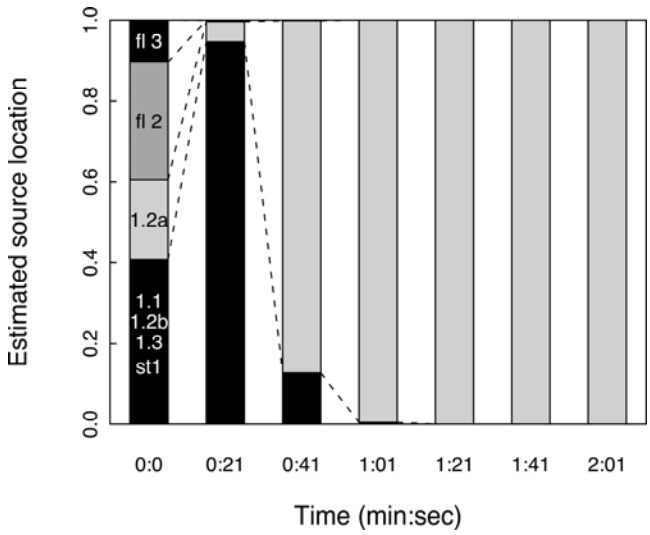

Figure 5. Probability of source being in location indicated, as estimated with the Bayes Monte-Carlo algorithm using threshold data with response time of 20 s, threshold level of 2.3%, and without added error. The actual release location is Room 1.2a. Time is referenced to the instantaneous release event.